Integrating Genes Affecting Coronary Artery Disease in Functional Networks by Multi-OMICs Approach

Baiba Vilne1,2 and Heribert Schunkert1,2*

1 Deutsches Herzzentrum München, Klinik für Herz- und Kreislauferkrankungen, Technische Universität München, Munich, Germany; 2 Munich Heart Alliance, German Centre for Cardiovascular Research, Munich, Germany

Coronary artery disease (CAD) and myocardial infarction (MI) remain among the leading causes of mortality worldwide, urgently demanding a better understanding of disease etiology, and more efficient therapeutic strategies. Genetic predisposition as well as the environment and lifestyle are thought to contribute to disease risk. It is likely that non-linear and complex interactions occur between these multiple factors, involving simultaneous pathological changes in diverse cell types, tissues, and organs, at multiple molecular levels. Recent technological advances have exponentially expanded the breadth of available -omics data, from genome, epigenome, transcriptome, proteome, metabolome to even the microbiome. Integration of multiple layers of information across several -omics domains, i.e., the so-called multi-omics approach, currently holds the promise as a path toward precision medicine. Indeed, a more meaningful interpretation of genotype-phenotype relationships and the development of successful therapeutics tailored to individual patients are urgently needed. In this review, we will summarize recent findings and applications of integrative multi-omics in elucidating the etiology of CAD/MI; with a special focus on established disease susceptibility loci sequentially identified in genome-wide association studies (GWAS) over the last 10 years. Moreover, in addition to the autosomal genome, we will also consider the genetic variation in our “second genome”—the mitochondrial genome. Finally, we will summarize the current challenges in the field and point to future research directions required in order to successfully and effectively apply these approaches for precision medicine.

Keywords: cardiovascular disease, multi-omics, genomics, transcriptomics, metabolomics, gut microbiome

INTRODUCTION

In the current era of high-potency statin therapy it becomes increasingly clear that even individuals with normal LDL-cholesterol levels without any conventional risk factors may develop atherosclerosis (1). The most pertinent manifestation of atherosclerosis is coronary artery disease (CAD), a highly complex disease, influenced by both multiple genetic risk variants and lifetime exposure to an atherogenic environment (2). A better understanding of the etiology of CAD and directions toward hitherto therapeutically not addressed disease mechanisms are urgently demanded (3). During the last 10 years, the genetic risk has been thoroughly explored in numerous genome-wide association studies (GWAS), leading to identification of >300 chromosomal loci which all significantly affect the risk of CAD (4–15). More than 90% of these common disease
risk variants are located outside the protein-coding regions and have modest effect sizes (2, 16). Collectively they explain only ~25% of the overall disease heritability. This suggests that genetic variation may contribute to disease risk in a non-linear, interactive and complex way (17), leading to pathological changes in diverse cell types, tissues, and organs, at multiple molecular levels (18).

Recent technological advances have exponentially expanded the breadth of available -omics data (17). High-throughput monitoring of the abundance of various biological molecules and determination of their variation between different conditions on a global scale has become possible, promoting a paradigm shift in the way we approach biomedical problems (19). At the same time, it has been increasingly recognized that no single type of data can fully capture the intricacy of most complex molecular traits that manifest collectively as disease phenotypes (20–22). Rather, it is the integration of multiple layers of information across several -omics domains, i.e., the so-called multi-omics approach [also referred to as integromics or panomics (19)], that holds the promise for precision medicine (Figure 1) (19).

Of note, integrative analysis across multiple-omics layers can be conducted in two ways (Figure 2): pair-wise data integration and multi-dimensional i.e., network-based integration (22). Furthermore, pair-wise integrations can be divided into genetic and non-genetic correlations (22). In the first case, DNA variants (i.e., allelic distributions of single-nucleotide polymorphisms; SNPs) are tested for association with down-stream omics markers such as transcriptomic alterations, protein, metabolite or methylation levels or quantitative and qualitative measures of microbiome, via the so called quantitative trait loci (QTL) mapping. In the second scenario, one would explore correlations between down-stream omics data, e.g., correlation of CpG methylation levels to transcript expression or between metabolome and gut microbiome, however it may be difficult to infer causal relationships in such case (22). Considering the largely unexplored role of the established CAD risk loci from GWAS (23) and the central dogma that genetic variations control the transcriptome, which in turn affects e.g., the proteome (20), and metabolome (Figure 2, middle panel), our main focus will be pair-wise integrations linking genetic variation related to CAD risk to other down-stream omics layers such as epigenome, transcriptome, proteome or metabolome. Although multi-dimensional integrations have been widely used in the field of cancer research, their application in the context of CAD has so far been limited (22). Moreover, in addition to the autosomal genome, we will also consider the genetic variation in our “second genome”—the mitochondrial genome and its contribution to CAD.

INTEGRATING GENETIC VARIATION AND EPIGENOME

Epigenomic signatures reflect various DNA modifications and may affect gene regulatory mechanisms that do not involve changes in the DNA sequence per se. Thereby, epigenomics may become a critical mediator of environmental influences and risk factors acting on the genome (20, 24). Three unique, but highly interrelated, epigenetic processes can be distinguished: DNA methylation, histone modifications (e.g., methylation, acetylation, phosphorylation, DP-ribosylation, and ubiquitination) and RNA-based mechanisms (e.g., microRNAs, long non-coding RNAs or IncRNAs, small interfering RNAs) (20, 24). Although, technically non-coding RNAs belong to the epigenome (20), we will discuss them in the next section, as the respective omics data are acquired via transcriptome profiling (RNA-seq).

DNA methylation and histone modifications are the best understood of the epigenetic mechanisms thus far and have been widely suggested to regulate gene expression and affect CAD risk factors including atherosclerosis, inflammation, hypertension and diabetes (25). DNA methylation consists of the covalent methylation of the C5 position of cytosine residues, when they are followed by guanine residues (CpG dinucleotides). It is partly heritable but it is also a dynamic process related to environmental stimuli and life style factors (26). Hedman et al. (27) analyzed epigenetic changes associated with lipid concentrations and identified a number of meQTLs, enriched in signals from GWAS on lipid levels and CAD. For example, genome-wide significant variants (rs563290 and its proxies), associated with LDL cholesterol and CAD at APOB, were meQTLs for a LDL cholesterol-related differentially methylated locus (Table 1 and Figure 3).

Furthermore, the CDH13 (T-cadherin) locus may present an interesting example in the context of epigenetics and CAD. Putku et al. (39) reported several genetic variants in the promoter of CDH13 as meQTLs in hypertension patients (Table 1), several of them being also associated with high molecular weight adiponectin, a known ligand for CDH13, the binding of which results in increased proliferation and migration of endothelial cells (39). Moreover, recently Nelson et al. (13) identified a genetic variant in the intron of CDH13, which affects expression of this gene in vascular tissues, and is genome-wide significantly associated with CAD (28) (Table 1). Interestingly, the expression levels of CDH13 and IncRNAs from the same locus showed positive correlations, suggesting a functional link, as IncRNAs are known to display correlations with the expression of their neighboring protein-coding target genes (48).

An exciting field of future research will be studies conducting parallel profiling of genetic variation with histone modifications and Hi-C and ChIA-PET-based chromatin contact maps to uncover local and distal histone quantitative trait loci (hQTLs) (49) in CAD patients.

Overall, considering the critical role of epigenetic modifications as a critical mediator of environmental influences on the genome (20, 24), we urgently need more investigations studying DNA methylation and other epigenetic modifications genome-wide and in large enough cohorts, ideally also elucidating the differences between tissues and cells in healthy vs. CAD patients. Moreover, this should be supplemented with careful documentation of multiple environmental and lifestyle factors over time, i.e., the envirome, as well as comprehensive clinical information to draw a link between the environment and CAD.
Multi-OMICs Networks in Coronary Artery Disease

**FIGURE 1** | Multi-omics approach for precision medicine. Multi-omics (i.e., genome, epigenome, transcriptome, proteome, metabolome, microbiome, and envirome) data are collected from patients and integrated to create their individual molecular signatures (i.e., complex biomarkers), which are then used to select an appropriate drug for a particular patient, thus improving the treatment efficiency and reducing the possible side effects.

**FIGURE 2** | Multi-omics (i.e., autosomal and mitochondrial genome, epigenome, transcriptome, proteome, metabolome, microbiome, and envirome) data integration can be conducted in two ways: pair-wise integrations, which can be further divided into non-genetic (left panel) and genetic correlations (middle panel). In the first case, one would examine the correlation patterns between the down-stream omics layers (e.g., metabolome and gut microbiome), whereas the second is achieved via the so called quantitative trait loci (QTL) mapping, linking genetic variation to methylation levels (meQTLs) or histone modifications (hQTLs), transcriptome (expression QTLs; eQTLs), protein (pQTLs), metabolite (mQTLs) or measures of microbiome (mbQTLs). Alternatively, multi-dimensional i.e., network-based integration approaches (right panel) exist, however their application in the context of CAD has so far been limited (22).

**INTEGRATING GENETIC VARIATION AND TRANSCRIPTOME**

Transcriptomics reflect genome-wide measures of RNA levels, both protein-coding RNA as well as the non-coding RNAs (i.e., microRNAs, IncRNAs, and small interfering RNAs) under specific conditions or in a specific cell. Moreover, the transcript levels are examined both qualitatively (i.e., which transcripts are present, identification of novel transcripts, splice sites, and RNA editing sites) and quantitatively (quantification of transcript abundance) (21).

**Protein-Coding RNAs**

Parallel assessments of genetic variation and transcriptome profiles across disease-relevant tissues, i.e., via mapping expression quantitative trait loci (eQTLs) to identify susceptibility genes (mainly protein-coding), has been the most commonly applied approach (28, 29, 50–52). Björkegren et al. have performed a number of integrative network analysis, linking CAD risk variants and transcriptome data in seven disease-relevant vascular and metabolic tissues, collected from up to 600 CAD patients during coronary artery bypass surgery (28, 29, 53, 54). From these investigations, visceral abdominal fat has emerged as an important gene-regulatory site for blood lipids. Several risk SNPs for HDL-, LDL-, and total cholesterol levels, as well as for CAD demonstrated significant eQTL effects in visceral abdominal fat (28, 29).

Huan et al. (30) also used integrative analysis to investigate the molecular mechanisms of blood pressure regulation and identified a blood pressure associated SNP (rs3184504) in SH2B3, also associated with the expression (eQTL) of several...
TABLE 1 | Genetic variation related to CAD/MI risk that has been associated with other down-stream omics layers such as transcriptome (mRNA, microRNAs and IncRNAs), epigenome, proteome or metabolome.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Tissue</th>
<th>Phenotypic Trait</th>
<th>SNP</th>
<th>Omics-Marker</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcriptome: mRNAs</td>
<td>Visceral abdominal fat</td>
<td>HDL cholesterol level</td>
<td>rs4148008</td>
<td>ABCAB/ABCA5</td>
<td>Franzén et al. (28)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total cholesterol level</td>
<td>rs11869286</td>
<td>STAR*D</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LDL cholesterol level</td>
<td>rs751557</td>
<td>EVS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs174546</td>
<td>TMEM258</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAD</td>
<td>rs12046679</td>
<td>PCSK9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Internal mammary artery, atherosclerotic aortic root</td>
<td>Blood pressure</td>
<td>rs920006</td>
<td>G3BP1</td>
<td>Foroughi Asl et al. (29)</td>
</tr>
<tr>
<td></td>
<td>Blood</td>
<td></td>
<td>rs6908994</td>
<td>PSORSIC3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs9930418</td>
<td>FLYWCH1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Circulating leukocytes, human coronary artery smooth muscle cells (HCASMC)</td>
<td>CAD</td>
<td>rs12190287</td>
<td>miR-224: TCF21</td>
<td>Miller et al. (31) Bastami et al. (32)</td>
</tr>
<tr>
<td></td>
<td>The effect of diet on plasma lipid levels</td>
<td>rs13702</td>
<td>miR-410:LPL</td>
<td>Richardson et al. (33)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAD</td>
<td>rs989727(rs7808424)</td>
<td>miR-202-5p:ASZ1</td>
<td>Bastami et al. (34)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs41269915(rs2229238)</td>
<td>miR-485-5p:UBE2Q1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs1556</td>
<td>hsa-miR-130a-5p:UBE2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs3088442</td>
<td>hsa-miR-130a-5p:SLC22A3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs2266788</td>
<td>hsa-miR-4722-5p:APOA5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs72932707</td>
<td>hsa-miR-4722-5p:CA1L</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HDL,LDL, and total cholesterol, triglycerides</td>
<td>rs2370747(rs7115089)</td>
<td>miR-100-5p,miR-125b-5p</td>
<td>Huan et al. (35)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAD</td>
<td>rs11042699 (rs6578985)</td>
<td>miR-483-3p:KIF2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Platelet count</td>
<td>rs4905998 (rs7149242)</td>
<td>miR-127-3P,miR-137,miR-370,miR-578a-5p,miR-582-5p,miR-431-5p,miR-433,miR-329,miR-409-3p,miR-494,miR-411-3p,miR-654-5p,miR-668,miR-543,miR-323a-3p,miR-337-3p</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcriptome: lncRNAs</td>
<td>Internal mammary artery, atherosclerotic aortic root</td>
<td>3p/5p ratio</td>
<td>rs13064131</td>
<td>miR-28:LPP</td>
<td>Civelek et al. (38)</td>
</tr>
<tr>
<td></td>
<td>CAD</td>
<td>rs1333045</td>
<td>FPKM1_group_33469_transcript_1</td>
<td>Ballantyne et al. (37)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>rs1333049</td>
<td>FPKM1_group_33469_transcript_2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2D</td>
<td>rs2383208</td>
<td>FPKM1_group_33469_transcript_6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epigenome:DNA methylation</td>
<td>Hypertension</td>
<td>rs113460654, rs124438378, rs12444338, rs6204565, rs8060301</td>
<td>ANRIL</td>
<td>McPherson et al. (38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Diastolic blood pressure, serum high-density lipoprotein, high molecular weight adiponectin</td>
<td>rs8060301</td>
<td>CDH13</td>
<td>Putk et al. (39)</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Data type</th>
<th>Tissue</th>
<th>Phenotypic Trait</th>
<th>SNP</th>
<th>Omics-Marker</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteome</td>
<td>Blood plasma</td>
<td>CAD</td>
<td>rs12740374, rs867186, rs1050362</td>
<td>Granulin (CELSR2/SORT1)</td>
<td>Chen et al. (40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Protein C (FROCR)</td>
<td>Howson et al. (14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>apoipoprotein L1 (DHX38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolome</td>
<td>Blood plasma</td>
<td>CAD</td>
<td>rs715</td>
<td>CPS1, urea cycle metabolites, plasma glycine</td>
<td>Hartiala et al. (41)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blood plasma</td>
<td></td>
<td>rs10450989(USP3), rs222613(HERC1), rs930491, rs11827377(STAT1), rs5853422(SEL1L), rs1969075(FBXO25), rs9591507, rs17573278, rs894840, rs9285184(SUGT1)</td>
<td>Circulating short-chain dicarboxylcarnitine (SC-DA)</td>
<td>Kraus et al. (42)</td>
</tr>
<tr>
<td>Multi-OMICS</td>
<td>Low HDL and inflammatory pathways</td>
<td></td>
<td>rs241437</td>
<td>TAP2</td>
<td>Laurila et al. (43)</td>
</tr>
<tr>
<td>Mitochondrial Genome</td>
<td>Blood</td>
<td>Hypertension</td>
<td>m.8701A&gt;G</td>
<td>HLA-DRB1, HLA-DQA1</td>
<td>Zhi et al. (44)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MT-ATP6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1** Continued

<table>
<thead>
<tr>
<th>Data type</th>
<th>Tissue</th>
<th>Phenotypic Trait</th>
<th>SNP</th>
<th>Omics-Marker</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

much less investigated are non-coding RNA transcripts, such as micro-RNAs (miRNAs) and long non-coding RNAs (lncRNAs). Recent evidence suggests that at least some of these may play a role in CAD (53–58). Although, technically non-coding RNAs belong to the epigenome (20), we will discuss them in this section, as the respective omics data are acquired via transcriptome profiling (RNA-seq).

**Micro RNAs**

miRNAs are involved in the transcriptional control of all main cell types participating in atherosclerosis progression, including endothelial cells, vascular smooth muscle cells, and macrophages (32, 59). Several studies have investigated the differential expression patterns of miRNAs in plasma/serum, microparticles, whole blood, platelets, blood mononuclear intimal, and endothelial progenitor cells in CAD vs. non-CAD patients, as summarized by Malik et al. (60). In majority of cases, up-regulation of different miRNA in CAD patients was observed (60). Moreover, growing body of evidence suggests that genetic variations in the miRNA targetome may lead to major deleterious outcomes (61, 62). For example, Miller et al. (31) have shown that an established CAD risk variant (rs12190287) resides in the 3’ untranslated region of a transcription factor TCF21 and alters the seed binding sequence for miR-224. Moreover, allelic imbalance studies in circulating leukocytes and human coronary artery smooth muscle cells have demonstrated a significant imbalance of the TCF21 transcript levels, which correlated with genotype at rs12190287, consistent with this variant contributing to allele-specific expression differences (31). Richardson et al. (33) have reported that a variant (rs13702) in the 3′-UTR of lipoprotein lipase (LPL) disrupts the binding of miR-410 and modulates the effect of diet on plasma lipid levels (33). Recently, Bastami et al. (34) performed a more systematic computational screening, by mapping the established CAD risk variants to the miRNA targetome, identifying several links between SNPs and miRNAs (Table 1; https://www.ebi.ac.uk/gwas/). In a recent study from our group (16), we also mapped CAD risk variants from the CARDIoGRAMplusC4D GWAS meta-analyses (9), to 3′ UTR regions of genes to assess their overlaps with predicted target miRNA binding sites. Interestingly, the 3′ UTR region of MRAS was predicted to be targeted by 29 miRNAs and 23 miRNAs were predicted to bind more than one candidate CAD gene (Table 1). Thus far, there have been relatively few studies investigating genome-wide miRNA eQTLs (miR-eQTLs). Huan et al. (35) identified a
Hedman et al. (27) identified SNP (rs515135) in an intron of APOB to be associated with LDL-C. Its proxy was also associated with CAD. Interestingly, this SNP represents a cis-meQTL. Black arrows indicate association findings. Red arrows indicate the presumed functional cascade leading to CAD.

Huan et al. (30) uncovered a blood pressure associated SNP (rs3184504) in SH2B3, which also associates with the expression (eQTL) of several genes, including SH2B3 itself, in the genetically inferred causal blood pressure gene sets. Rs3184504 has been previously also associated with CAD risk. (9) Black arrows indicate association findings. Red arrows indicate the presumed functional cascade leading to CAD.

genetic variant (rs2370747) associated with miR-100-5p and miR-125b-5p expression, a proxy SNP of which was also associated with lipid traits (HDL-, LDL-, and total cholesterol as well as triglycerides). Moreover, it was found that both miRNAs were also differentially expressed in relation to HDL cholesterol (35).

Civelek et al. (36) examined the genetic regulation of human adipose miRNA expression and its consequences for metabolic traits. Interestingly, this study showed, how genetic variation might influence the processing of miRNAs, i.e., the ratio of miRNA expression from the 3p and 5p arms. It is known that a miRNA precursor can give rise to two mature miRNAs from the 3p and 5p arm, one of which usually having higher expression than the other. The 3p/5p ratios of several miRNAs have been shown to be significantly different among various healthy tissues (63) and altered in pathological conditions compared with healthy controls (64). Civelek et al. demonstrated a significant association of the SNP rs13064131 with the 3p/5p ratio of miR-28, encoded from the LPP gene (Figure 5) (36). However, the SNP was not associated with the expression levels of the LPP transcript itself or with the abundance of miR-28-3p or miR-28-5p, suggesting that its effect on the 3p/5p ratio may be independent of transcription, possibly via degradation or stabilization mechanisms.

Long Non-Coding RNAs

The recent discovery of an extensive catalog of lncRNAs—i.e., long RNA transcripts that do not code for proteins—has opened a new perspective on the importance of the RNA-based mechanisms in gene regulation (24). LncRNAs are emerging as important regulators of various cellular processes, with many possible implications in cardiovascular disease pathophysiology (57, 58). In fact, the most prominent CAD risk locus at Chr9p21 (66, 67) harbors the lncRNA—ANRIL (Antisense Non-coding RNA in the INK4 Locus, CDKN2B antisense RNA). From these, rs10757274 is the strongest genetic predictor of early MI and is not associated with established CAD risk factors such as lipoproteins or hypertension, making ANRIL a key candidate (38). Interestingly, ANRIL is found both as a linear lncRNA (linANRIL), the transcript levels of which are known to positively correlate with disease severity (68), and is also capable of forming RNA circles (circANRIL).
Recently, Holdt et al. (69) demonstrated that circANRIL regulates the maturation of precursor ribosomal RNA (pre-rRNA), by this impairing ribosome biogenesis and inducing nucleolar stress and apoptosis in vascular smooth muscle cells and macrophages (Figure 6). Carriers of the CAD-protective haplotype at 9p21 showed significantly increased expression of circANRIL (69).

Currently, there have not been many large-scale studies on lncRNAs in the context of CAD, though. Ballantyne et al. (37) recently conducted a genome-wide interrogation of long intergenic non-coding RNAs (lincRNAs) that associate with cardiometabolic traits in GWAS, including CAD and also identified a number of CAD/MI and type 2 diabetes associated SNPs at Chr9p21 that overlapped lincRNA transcripts (Table 1) (37). In STARNET (28), 5.4% of the identified cis-expression quantitative trait loci (eQTLs) were related to the expression of lncRNAs, however these have not been further explored, so far. Overall, more studies focusing on non-coding RNAs in different CAD relevant tissues in large enough cohorts will be required to yield insights into the possible functional roles of this portion of transcriptome and its genetic determinants, in healthy and disease states. Moreover, considering that lncRNAs are generally found to be more lowly-expressed, sufficient depth of coverage for RNA-seq experiments will need to be guaranteed (21).

**INTEGRATING GENETIC VARIATION AND PROTEOME**

Proteomics uses high-throughput approaches (mainly MS-based) to quantify protein abundance, post-translational modifications and interactions (e.g., using phage display and yeast two-hybrid assays) in a tissue, cell or fluid compartment, such as plasma or urine (21). Considering that the transcriptome is not linearly proportional to proteome, that proteins are the biomolecules that execute cellular functions, and that many human diseases ultimately result from alterations in the proteome (70), such studies are urgently needed to facilitate the explorations of CAD etiology. However, proteome studies are still rare in relation to CAD, mostly due to the complex methodology involved. There have been some investigations in the past few years, aiming at characterizing the proteomes of several CAD-related tissues and cell types, including human arterial smooth muscle cells (71), platelets (72), as well as body fluids such as urine (73).

Only few studies (14, 40) have analyzed genetic variants that modify protein levels, i.e., the so-called protein quantitative trait loci (pQTLs) (Table 1). Chen et al. (40) assayed a pre-selected set of plasma proteins, identifying several pQTLs that overlapped with CAD risk SNPs and also explained a substantial proportion of inter-individual variation in protein abundance. For example, rs12740374 at the CELSR2/SORT1 locus, a variant associated with lipids and CAD, explained 15% of inter-individual variation in plasma granulin levels (Figure 7). Interestingly, progranulin binds to SORT1 and Sort1 knockout mice show markedly elevated levels of progranulin (40). Recently, it was also demonstrated that progranulin is involved in lysosomal homeostasis and lipid metabolism (74).

As the proteomics technologies improve over time (21), more genome-wide investigations of CAD-related alterations in proteome and also phosphoproteome in increasing numbers of disease relevant tissues are expected to be conducted in the near future. However, as proteins are more sensitive to their environment (21), caution will have to be taken during sample preparation steps to obtain accurate and reproducible results.

**INTEGRATING GENETIC VARIATION AND METABOLOME**

An important additional functional layer in mutli-omics data integration is the metabolome, as it represents an integrated state of all genetic, epigenetic and environmental factors, thus providing a link between genotype and phenotype (75). Metabolomics is an omics field that systematically identifies and quantifies multiple small molecule (typically <1,500 Daltons) types, such as amino acids, fatty acids, carbohydrates and biochemical intermediates, i.e., metabolites (21). A plethora of metabolites in blood and urine have been associated with CAD and subsequent cardiovascular events (76–79) and have been
rs12740374 at the CELSR2/SORT1 locus, (40) a variant associated with lipids and CAD, was recently found to display pQTL effects on plasma granulin levels, and pro-granulin is known to bind to SORT1. More recently, it was also demonstrated that progranulin is involved in lysosomal homeostasis and lipid metabolism (74). Demonstrated as promising biomarkers discriminating CAD vs. non-CAD subjects (78), as well as between thrombotic MI and stable CAD cases (80). Kraus et al. (42) recently identified several genetic loci demonstrating associations with blood plasma metabolites (i.e., metabolomic quantitative trait loci; mQTLs), the strongest findings being for the circulating short-chain dicarboxylacylcarnitine (SCDA) metabolite levels with variants in genes that regulate components of endoplasmic reticulum (ER) stress (Table 1 and Figure 8) (42).

Besides blood and urine, metabolomic profiles of vascular and metabolomic tissues such as subcutaneous fat will need to be generated, ideally in conjunction with other omics layer data. Especially, gut microbiome would be of utmost interest, considering the close link between the two (81).

However, of note, metabolic profiles are even more prone to variability affected by sample preparation and storage conditions, as well as by several other factors including patient heterogeneity (21). Hence, the required sample size has to be carefully considered, to inspire confidence in the generated results.

**INTEGRATING GENETIC VARIATION AND MICROBIOME**

Microbiomics investigates all the microorganisms of a given community, including bacteria, viruses, and fungi, collectively known as the microbiota (and their genes constituting the microbiome) (21). The human microbiome is enormously complex and there are substantial variations in microbiota composition between individuals resulting from seed during birth and development, diet and other environmental factors, drugs and age (21). Thousands of different bacterial species make up the human microbiomes, from which there is a small number of abundant species and a large number of rare or low abundance species, the differential functions of which remain poorly understood (82). Currently, several large scale initiatives are emerging including the American Gut Project http://americangut.org/ and the British Gut Project http://britishgut.org/, which are expected to produce a rich collection of anonymised human gut samples and lifestyle information for medical researchers.

Gut microbiome has emerged as another rich source of information and as a possible new player contributing to the CAD/MI pathogenesis (82–84). It has long been known that bacteria activate inflammatory pathways, and recent data demonstrate that the gut microbiome may also affect lipid metabolism and influences the development of obesity and atherosclerosis (84), suggesting that gut microbiota could be used as a diagnostic marker for CAD (85). The most investigated is the association between gut microbiota and fasting plasma trimethylamine N-oxide (TMAO) levels, a gut microbiota-dependent metabolite, previously also associated with CAD and stroke (81, 86). Org et al. (81) demonstrated that certain blood plasma metabolites strongly correlated with gut microbial community structure and that some of these correlations may be specific for the pre-diabetic state. LeChatelier et al. (84) used quantitative gut microbiome information to distinguish between individuals with “high bacterial richness” and “low bacterial richness,” were the latter were characterized by increased adiposity, insulin resistance and dyslipidemia in addition to a more pronounced inflammatory phenotype. Le Chatelier Fu et al.
INTEGRATING GENETIC VARIATION AND MULTIPLE OMICS DATASETS

An integrative analysis of genetic variation and transcriptome with additional high-throughput measurements may greatly improve the predictive power of disease networks. Zhu et al. (89) However, the number of studies conducting multi-omics integrations in the context of CAD is limited so far. Miller et al. (90) integrated genetic variation with investigations of chromatin state, enhancer activity and TF binding in human coronary artery smooth muscle cells and demonstrated, for example, that one of the lead candidate variants, rs17293632, located within an intergenic region of the SMAD3 gene, overlaps an open chromatin region. Moreover, it was observed that the major risk C allele was more associated with open chromatin and resided in a canonical AP-1 motif, which was effectively destroyed by the minor protective T allele. Preferential AP-1 binding to the risk C allele was experimentally validated using allele-specific ChiP analyses. Miller et al. (90) and Kraus et al. (42) performed a pathway-level integrative analyses, linking genetics, epigenetics, transcriptomics, and metabolomics profiles and implicating the ubiquitin proteasome system in cardiovascular disease pathogenesis. This study observed associations of circulating short-chain dicarboxylacylcarnitine (SCDA) with variants in ER stress genes, whereof several genetic variants (Table 1 and Figure 8) in FBXO25 and SUGT1 genes also demonstrated evidence of cis-regulation in expression quantitative trait loci (eQTL) analyses and independently predicted CAD events (42). Moreover, two other genes from the same ER stress pathway—BRSK2 and HOOK2—were identified as differentially methylated, when comparing individuals with high and low SCDA levels. Subsequently, experimental validation using culture of human kidney cells in the presence of levels of fatty acids found in individuals with cardiometabolic disease, demonstrated induced accumulation of SCDA metabolites in parallel with increases in the ER stress marker BiP (42).

Shu et al. (20) investigated shared genetic regulatory networks for CAD and type 2 diabetes (T2D) and their key intervening drivers in multiple populations of diverse ethnicities by performing an integrative analysis of five multi-ethnic GWAS for CAD and T2D, eQTLs, ENCODE, as well as tissue-specific gene network models (both co-expression and graphical models) from disease-relevant tissues. This study identified pathways regulating the metabolism of lipids, glucose and branched-chain amino acids, as well as pathways governing oxidation, extracellular matrix and immune response as shared pathogenic processes for both diseases and identified 15 key drivers including HMGCR, CAV1, IGF1, and PCOLCE, whose network neighbors collectively accounted for ~35% of known GWAS hits for CAD and 22% for T2D (20). Laurila et al. (43) applied a combined approach using both QTLs and canonical pathway analysis to link genomics and transcriptome analysis from the subcutaneous adipose tissue and plasma HDL lipidomics profiling, highlighting change in HDL particle quality toward putatively more inflammatory and less atheroprotective phenotype in subjects with low HDL, due to their reduced antioxidative capacity. Within the HLA region, this study found two significant, dose-dependent cis-eQTL associations with low HDL and inflammatory pathways: rs241437 in the intron of TAP2 and rs9272143 between HLA-DRB1 and HLA-DQA1, the latter also being associated with down-regulation of antioxidative pathways in HDL particles (43).

The application of multi-omics integrations in the field of CAD has so far been limited (22). Obviously, one of the main reasons for this is the current lack of appropriate data in large enough cohorts. However, considering the great promise such studies hold for precision medicine, it is expected that parallel measurements on multiple omics layers will be rapidly collected during the next couple of years, allowing also a comprehensive comparison, validation and improvement of the existing computational integration methods.

MITochondrial GENetic VARIation and DOWNstream OMICS DATASETS

Dysfunction of mitochondria has been increasingly associated with obesity-related cardiometabolic diseases and CAD (91). Thus, genetic variation in the mitochondrial DNA (mtDNA), which codes for the 37 OXPHOS genes as well as further >1000 nuclear-coded genes imported into mitochondria constituting essential components for their proper functioning, needs exploration for a better understanding of CAD genetics. The mitochondrial haplogroup T (45) and mtDNA variants m.16189T>C (46) and m.15927G>A (47) have been associated with CAD in different ethnic groups. Another mitochondrial variant, m.8701A>G, has been associated with hypertension (44). This variant is located in MT-ATP6 (ATP synthase/complex V F0 subunit 6) gene, which is part of the ATP synthase enzyme, responsible for the final step of oxidative phosphorylation, and, on the functional level, using mitochondrial hybrid cells (cybrids), it has been shown that it alters mitochondrial matrix pH and intracellular calcium dynamics (Figure 9) (92).

Similarly, other mitochondria-related omics data investigations could be of interest in the context of CAD, as Baccarelli et al. (93) reported that ATP synthesis genes including protein-encoding cytochrome c oxidase genes (MT-CO1, MT-CO2, and MT-CO3) and MT-TL1 were hypermethylated in platelets of CAD cases as compared to healthy controls (93). Using eQTLs in seven CAD relevant vascular and metabolic tissues (53) in conjunction with established CAD risk loci from GWAS (9) and time-resolved transcriptome data in the aortic arch in mice with reversible hypercholesterolemia (94, 95) we
recently demonstrated a massive down-regulation of nuclear-encoded mitochondrial genes (96), specifically at the time of rapid atherosclerotic lesion expansion and foam cell formation, which was largely reversible by genetically lowering plasma cholesterol. Both mitochondrial signature genes were supported as causal for CAD in humans, as eQTLs representing their genes significantly overlapped with disease risk SNPs. In line with this, the STARNET (28) study recently examined mitochondrial (i.e., mtDNA-derived) gene expression and a markedly lower expression of mitochondrial genes in the atherosclerotic aortic arterial wall as compared to non-atherosclerotic arterial wall.

Furthermore, genetic variation of mitochondrial metabolome has remained largely unexplored. Hartiala et al. (41) searched for genetic factors associated with plasma betaine levels and determined their effect on CAD risk. This resulted in the identification of two significantly associated loci on chromosomes 2q34 and 5q14.1. The lead variant on 2q24—rs715—localized to carbamoyl-phosphate synthase 1 (CPS1), which encodes a mitochondrial enzyme that catalyzes the first committed reaction and rate-limiting step in the urea cycle. Rs715 was also significantly associated with decreased levels of urea cycle metabolites and increased plasma glycine levels. Finally, rs715 yielded a strikingly significant and protective association with decreased risk of CAD in women (41).

Finally, in recent years, it has become increasingly evident that the gut microbiome produces metabolites that influence mitochondrial function and biogenesis (97), hence the ancestral gut microbiome-mitochondrion connection and its relation to CAD might need to be explored in the near future, as well.

Recent progress in next-generation sequencing (NGS) techniques has set a scene for a second “gold rush” in mitochondrial genomics and mtDNAs are presently the most sequenced type of eukaryotic chromosome (98). At the same time, multi-omics investigations in mitochondria, mapping the genomes, transcriptomes, proteomes, and metabolomes in parallel, apart from yeast (99) have not been conducted yet. Hence, although, mitochondrial dysfunction has been associated with many human diseases, the respective proteins and pathways are not well-characterized (99), presenting an exciting future field of investigation, especially considering the fact that mitochondria play a key role in plasticity and adaptation to environmental change, including adaptation to physiological stress (100).

CONCLUSIONS AND FUTURE DIRECTIONS

Given that CAD like other common complex disorders develops over time and involves both genetics and environment, full mechanistic insight will require coordinated sets of several-omics data at multiple time points, collected from many disease relevant tissues and body fluids in large enough cohorts (20, 21). Environmental risk factors can interact with the genome and perturb the epigenome to further modulate the transcriptome and proteome (20). Therefore, comprehensive monitoring and careful documentation of multiple environmental and lifestyle factors over time, i.e., the envirome, will be indispensable to yield significant insights into the complex etiology of CAD. Moreover, imaging and electronic health record data also will need to be considered. As more-omics and other data are generated, novel methods for efficient data integration, modeling, visualization and interpretation will be urgently needed to efficiently cope with this multi-dimensional data (101), and translate it into actionable precision medicine tools. Although, there has been major progresses in the development of multidimensional data integration algorithms and tools, the field is still in its infancy and the flexibility, effectiveness and robustness of data integration to extract biological insights is still restricted, especially when clinical outcomes (e.g., stable CAD vs. MI) need to be modeled (22, 101). In addition we still face a number of technical challenges related to patient sampling and profiling. For example, as already recognized by Hasin et al. and others (20, 21) human studies are often affected by various confounding factors, which are difficult or even impossible to control for (e.g., diet and medications). Clearly, also the available sample size will play an important role for the multi-omics approach to produce meaningful insights into CAD (21) and allow the generation of reliable prediction models for more efficient design of therapeutics, tailored to individual needs. According to Hasin et al. an underpowered study may not only miss true signals, but is also more likely to produce false positive results (21). Furthermore, already before and during data collection, careful
attention has to be paid to data analysis requirements, e.g., sufficient depth of coverage for RNA-seq experiments (21).

**AUTHOR CONTRIBUTIONS**

BV and HS drafted and edited the manuscript.

**FUNDING**

This work was supported by grants from the Fondation Leducq [CADgenomics, 12CVD02], the German Federal Ministry of Education and Research (BMBF) within the framework of ERA-NET on Cardiovascular Disease, Joint Transnational Call 2017 [ERA-CVD: grant JTC2017_21-040], within the framework of target validation [BlockCAD: 16GW0198K], within the framework of the e:Med research and funding concept [AbCD-Net: grant 01ZX1706C and e:AtheroSysMed: grant 01ZX1313A-2014], and the European Union Seventh Framework Programme FP7/2007–2013, under grant agreement no. HEALTH-F2-2013-601456 (CVgenes-at-target). Further grants were received from the Deutsche Forschungsgemeinschaft (DFG) as part of the Sonderforschungsbereich CRC 1123 (B2).

**REFERENCES**

27. Hedman ÅK, Mendelson MM, Marioni RE, Gustafsson S, Joehanes R, Irvin M.R. Epigenetic patterns in blood associated with lipid traits predict incident coronary heart disease events and are enriched for results...


**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Vilne and Schunkert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.